STRETCHR: THE INTELLIGENT DATASTACK IMPACT ON LOG STRATEGY

Note: This paper assumes the reader is familiar with Stretchr's core feature set and architecture, an overview of which is outlined in the paper, "An Introduction to Stretchr: The Intelligent Datastack"

TABLE OF CONTENTS

Contents

The Importance of Log Strategy

THE LOST POTENTIAL OF LOGS

Regardless of industry, maintaining logs for production environments is a valuable, and often required, addition to any infrastructure. Unfortunately, storing and maintaining logs in even a small environment can become a large undertaking as scaling, physical hardware requirements, maintenance and the collection of dissimilar data from multiple parts of the system must be considered.

Adding to these challenges is the lack of useful querying capabilities - the integration of the data into existing workflows most often requires substantial backend development, user/permissions systems development, database optimization, and determination of how to best collect and expose the data for the various interfaces that are used to mine insights from the data store. Creating applications that provide true log transparency are very often discouraged or avoided because of the time consuming nature and high development cost of the process. Instead, organizations commonly address such problems by employing a complex mosaic of third-party solutions and individual employees' own toolbox of scripts and undocumented techniques for isolating an application's issue.

In this situation system administrators and business owners are unlikely to have a true understanding into overall system activity and key drivers of system performance. As a result, many companies stop short of realizing the full potential of their logs. When performance goes wrong, discovering whether it is a network failure, a bug in the code of a sub-system, user error, or some unusual combination of circumstances, is very difficult and results in a highly complex, manual, and time consuming debugging exercise.

THE BENEFIT OF A HOLISTIC APPROACH TO LOGGING

With a holistic approach to auditing and monitoring, however, log data becomes an invaluable source of information that is fundamental to problem identification and core issue diagnosis like understanding system inefficiencies or performance bottlenecks in code. Log data paints a real usage picture of an enterprise's products and services. It provides key decision inputs regarding how and when to scale infrastructure, where to dedicate development resources to drive the most value for users, and when to schedule maintenance and support downtime. Managing and maintaining log data traditionally requires an ever increasing amount of storage, processing, indexing, and human resources.

LOG STRATEGY & THE INTELLIGENT DATASTACK

Stretchr simplifies application and system logs collection, scaling, query analytics and reporting to a point of triviality. It's schemaless capabilities allow it to collect, manage and analyze logs from any web enabled device while respecting each device's unique and varying data structures.

By acting as a centralized data store for all of this information, Stretchr quickly becomes the source of operational and customer systems activity information, while keeping infrastructure at the right scale, and ensuring interactions remain quick and stable. By doing so, Stretchr also extracts and increases the value of a system's log repository.

This paper discusses how Stretchr can be used to create a holistic log strategy, and how it is utilized to expose valuable insights into the health of a system. The strategy has four key components:

- Log Collection
- Log Scaling (up and down)
- Log Query, Analytics & Reporting
- Increasing the Value of the Log Repository

Log Collection

DISPARATE LOG COLLECTION PROVIDES THE FOUNDATION TO SYSTEM USAGE TRANSPARENCY

Every standard and internally developed application, and every device or utility, generates logs in its own format with a unique context and domain specific data.

For example, a single web request from a user will often touch many different services within a system. Every standard and internally developed application, and every device or utility, generates logs in its own format with a unique context and domain specific data. A web server logs the URL, time and duration of the request while a web application reports on the request’s progression through its own internal structures. The request may contain database interactions, calls to external API's or interact with a local file system. All of which will produce a diverse collection of log entries. And while the formatting and context of the entries will differ, each contributes to an overall understanding of users’ product and system utilization. This siloe’d nature of logs traditionally limits the ability to include data from all sources when making specific requests.

STRETCHR INGESTS LOG DATA IN ANY FORMAT OR DATA REPRESENTATION

Stretchr's flexible RESTful API allows any network enabled application to contribute log data in any format or data representation (e.g. JSON, XML, CSV etc.). Its optimistic design is indifferent to log entries with differing fields, or if some have fields that others don't. Applications can be written to communicate with Stretchr directly, or more likely, existing logs can be ingested periodically with lightweight data synchronization scripts. This flexibility enables the collection and utilization of dispersed logs with minimal development or interference with existing infrastructures.

Log Query, Analytics & Reporting

HAVING A WEALTH OF LOG DATA BECOMES TRULY USEFUL WHEN MINED, PARSED, AND UNDERSTOOD

Querying and reporting are vital to quickly finding relevant entries in response to a bug report, creating heuristic views of systems interactions, or understanding how customers are making use of the system's capabilities.

Stretchr's advanced querying capabilities and nested data structure allows for easily filtering of large amounts of log data rapidly, just by tweaking the URL of the requests. Rather than expending developer resources to build a home grown logging solution, or installing more rigid out of the box products, developers using Stretchr rapidly build meaningful and specific tools that address key log-related issues.

Its advanced querying capabilities allows for a filter, enabling such queries as:

- How many log entries have a `duration` field?
    - Find all the log entries that don't have one so I can calculate it from the `start` and `end` fields.
- Find all log entries that have any items in the `errors` array
- How many requests were made from a specific I.P. address
- How many times did we write to disk storage today?

- What is our rate of growth over the past month and where do we expect it to be at the end of next month?

Once a suitable set of data has been selected, integrating with existing analytics tools or developing a mashup dashboard of one's own, is often accomplished with a few lines of code.

Log Scaling

WITH STRETCHR, LOG SCALING BECOMES AS EASY AS SPINNING UP A MACHINE INSTANCE

In a logging environment, scaling infrastructure up and down is important to avoid potential data loss during periods of high activity or wasted expenses during periods of low activity.

With Stretchr, scaling is as simple as spinning up new API instances. Instances are registered with zero downtime, ensuring minimal impact to end users. This, along with its decoupled nature from the datastore gives you a powerful toolset to use when right-sizing your data stack deployment.

Stretchr's abstracted design allows it to run natively on either physical hardware or in a virtualized environment, providing flexible implementation options for any situation.

Increasing the Value of a Log Repository: Rapid Small Project Application Development

ONCE THE LOG DATA IS EXPOSED IT ENABLES RAPID, SMALL APPLICATION DEVELOPMENT

Logs are often trapped inside a text file within a system, spread across many systems with complex security policies and access controls, or stuck in a legacy relational database. Stretchr's RESTful API capability "frees" this data and exposes it in a secure, predictable fashion with powerful querying features that any network enabled application can use.

Once the data is exposed, development focus often changes from large scale applications for a wide variety of uses, to rapidly developed smaller applications that more intimately meet the needs of smaller user groups. Example applications that have been built off of Stretchr log data include:

- A dashboard that displays the overall activity across large systems including individual customer billing detail and graphics of the busiest machines in the network

- A bot that updates a sales counter on the marketing website every time there is a successful sale

- A bot that watches for errors in the logs, and flashes the lights in the office red when they occur.

Rapid, small project development is further enabled by Stretchr's powerful permissions system built right into the core. Controlling who has access to any piece of data is possible with just a few simple rules. For example, it is easy to specify rules to indicate that:

- Only registered users can access the log data

- Only administrators can see the `system` information in log entries

- Users of the system can view their own logs, but not each other's

Such data exposures would often be discouraged in a more traditional development environment, when accomplishing the level of security necessary to make it possible would often require weeks of backend development before the interfaces could even be started.

Conclusion

Creating a comprehensive log strategy offers powerful benefits to any organization, but is unfortunately a challenging task riddled with potential blockers. Implementing your strategy with Stretchr's intelligent data stack at its core dramatically simplifies the process of centralizing, utilizing and scaling your solution. Once centralized, log data can be securely exposed to any web enabled device with fine-grained access control, powerful querying capabilities and a predictable API - increasing developer productivity and allowing them to work more closely with product teams to rapidly release powerful internal tools and customer facing applications.

Company Information

Stretchr, Inc.
4209 Niblick Dr.
Longmont, CO 80503
www.stretchr.com
@stretchrcom